11-1-2008

The effect of clustering on the uncertainty of differential reflectivity measurements

A. R. Jameson
RJH Scientific, Inc., El Cajon, California

A. B. Kostinski
Michigan Technological University

Recommended Citation

Jameson, A. R., & Kostinski, A. B. (2008). The effect of clustering on the uncertainty of differential reflectivity measurements. *Journal of Applied Meteorology and Climatology, 47*(11), 2816-2827.
http://dx.doi.org/10.1175/2008JAMC1860.1
Retrieved from: https://digitalcommons.mtu.edu/physics-fp/219

Follow this and additional works at: https://digitalcommons.mtu.edu/physics-fp
Part of the Physics Commons

The Effect of Clustering on the Uncertainty of Differential Reflectivity Measurements

A. R. JAMESON

RJH Scientific, Inc., El Cajon, California

A. B. KOSTINSKI

Michigan Technological University, Houghton, Michigan

(Manuscript received 27 August 2007, in final form 11 March 2008)

ABSTRACT

One of the most important avenues of recent meteorological radar research is the application of polarization techniques to improve radar rainfall estimation. A keystone in many of these methods is the so-called differential reflectivity Z_{DR}, the ratio of the reflectivity factor Z_H at horizontal polarization backscattered from a horizontally polarized transmission to that corresponding to a vertically polarized transmission Z_V. For such quantitative applications, it is important to understand the statistical accuracy of observations of Z_{DR}. The underlying assumption of all past estimations of meteorological radar uncertainties is that the signals obey Rayleigh statistics. It is now evident, however, that as a radar scans, the meteorological conditions no longer always satisfy the requirements for Rayleigh statistics. In this work, Z_{DR} is reconsidered, but this time within the new framework of non-Rayleigh signal statistics. Using Monte Carlo experiments, it is found that clustering of the scatterers multiplies the standard deviation of Z_{DR} beyond what is always calculated assuming Rayleigh statistics. The magnitude of this enhancement depends on the magnitudes of the clustering index and of the cross correlation between Z_H and Z_V. Also, it does not depend upon the number of independent samples in an ensemble estimate. An example using real radar data in convective showers suggests that non-Rayleigh signal statistics should be taken into account in future implementations of polarization radar rainfall estimation techniques using Z_{DR}. At the very least, it is time to begin to document the prevalence and magnitude of the clustering index in a wide variety of meteorological conditions.

1. Introduction

For over 25 years, the applications of radar polarization measurements have been thoroughly explored by a multitude of investigators. One of the most important avenues of research has been that directed toward improving radar rainfall estimation. A keystone of many of these methods is the so-called differential reflectivity, the ratio of the reflectivity factor Z_H backscattered at horizontal polarization from a horizontally polarized transmission to that corresponding to a vertically polarized transmission Z_V. Indeed, it now appears that $Z_{\mathrm{DR}} = Z_H/Z_V$ is destined to play an increasingly important operational role when polarization technology is added to current and future national meteorological radar systems.

It is important, then, to understand the measurement accuracy of Z_{DR}, usually expressed in terms of the standard deviation. Consequently, it is not surprising that the signal statistics of Z_{DR} have been thoroughly treated in past studies beginning with Bringi et al. (1983). Recently, however, some investigators (Gorgucci et al. 2006, p. 3039) even claim "typical" accuracies of 0.2 dB (about 5%). While that would be wonderful if true, this work shows why that degree of accuracy is often unlikely, in reality. It is important, then, not to oversell the anticipated accuracy of some polarization rainfall estimation techniques that are likely to be implemented in future national meteorological radar systems, particularly in light of the inherent variability of the rainfall rate itself (Jameson 2008).

The underlying assumption of all past studies is that the signals obey Rayleigh statistics (Rayleigh 1877) as

Corresponding author address: A. R. Jameson, 5625 N. 32nd St., Arlington, VA 22207-1560.
E-mail: arjatrjhsci@earthlink.net

DOI: 10.1175/2008JAMC1860.1

reexpressed for radar meteorology by Marshall and Hitschfeld (1953). The derivation of Rayleigh statistics assumes that all waves from each particle scatter independently and that conditions are statistically stationary (or equivalently, spatially homogeneous). Thus, at any location, the net observed wave is then the sum of all the independent waves having randomly distributed phases.

Rather than directly measuring the amplitude and phase of the net wave, for radars it is more convenient to measure the so-called real $\mathcal{R}$ and imaginary Q orthogonal components. These can be thought of as the x and y axis projections, respectively, of a complex phasor (vector) rooted at an axis origin that has a magnitude equal to the amplitude of the net wave and an angular position around a circle given by the phase. From this perspective, Rayleigh statistics implies that the $\mathcal{R}$ and Q components are Gaussian distributed and are statistically independent. The statistical distributions of the amplitude (the Rayleigh distribution) and intensities (exponential distribution) can then be readily calculated.

As Marshall and Hitschfeld demonstrated, this approach is useful for a basic understanding of radar signals at one polarization or frequency. They also showed that mean intensity calculated from k statistically independent samples could then be written as an Erlang distribution having a shape dependent on k. The larger the k, the narrower and more Gaussian-like the distribution of the mean values. While as important as these findings are, there are situations when they do not apply.

For example, there have been past measurements suggesting that, at times, the statistics may deviate from Rayleigh (e.g., Schaffner et al. 1980). Unfortunately, the origins of the signals were not completely understood, leading most observers to conclude that such deviations were "unusual" and not a challenge to the predominance of Rayleigh statistics. We return to this in a moment.

While most radar measurements involve only one parameter, the complicated nature of meteorological conditions has forced increasing sophistication. For example, for some purposes it is often useful to measure two parameters that, when combined, yield greater information. One example is dual-wavelength hail detection involving the ratio of two radar reflectivity factors measured at two different frequencies. Such ratios present a statistical problem especially when they involve variables with probability density functions (pdfs) having extended tails. In the case of dual-wavelength hail detection, the statistics were first considered by Srivastava and Carbone (1971). Since the samples at the two frequencies are entirely statistically independent, the statistics of the dual-wavelength ratio is then the ratio of two Erlang distributions, which, it turns out, is a beta distribution, as shown in many texts on statistics. That distribution is usually asymmetric, often having quite extended tails that seriously challenge the statistical reliability of such measurements.

Indeed, if Z_H and Z_V were completely statistically independent, Z_{DR} would also be beta distributed, having long, extended tails, thereby seriously compromising the usefulness of Z_{DR}. This can be understood by simply realizing that when the two signals are statistically independent, a small value of Z_V might occur simultaneously with a large value of Z_H so that Z_{DR} could, at times, reach tremendous values. This would not necessarily be an infrequent event either. Fortunately, in precipitation, Z_H and Z_V are usually highly correlated ($\rho_{HV} \sim >0.98$). This correlation severely restricts the range of values of Z_{DR}. That is, large Z_H and Z_V as well as small Z_H and Z_V will occur simultaneously so that their ratio remains close to the expected value regardless of extreme fluctuations in either quantity. Consequently, Bringi et al. (1983) and subsequent studies all treat Z_{DR} as the ratio of two highly correlated quantities, each consistent with Rayleigh signal statistics.

Letting $\langle\ \rangle$ denote an ensemble average and E the expected value, the normalized differential reflectivity can be expressed as $X = (\langle Z_H\rangle/\langle Z_V\rangle)/[E(Z_H)/E(Z_V)]$. Using Eq. (3) in Bringi et al. (1983) and changing the variables from normalized amplitudes to normalized differential reflectivities, the pdf is given by

$$P(X) = \frac{(1 + X)(1 - \rho_{HV})}{[(1 + X)^2 - 4\rho_{HV}X]^{3/2}}, \tag{1}$$

where ρ_{HV} is the cross polarization between the backscattered intensities at the two polarizations. While this distribution exists and integrates to unity over $[0, \infty]$, and while there is a modal (expected) value, the mean and variance (or any other moment) do not exist because the tail of the distribution goes as $1/X^2$ (Fig. 1), as can be derived from (1) for large X, so that the moments do not converge. Bringi et al. (1983) circumvent this peculiarity by considering, instead, the statistics of an estimator of the expected value that comes from the summation of the Gaussian $\mathcal{R}$ and Q components at the two polarizations. The resulting statistics can then be treated using the analyses of Krishnaiah et al. (1963) for correlated bivariate chi distributed variables which arise from such summations. The Cramer–Rao lowest possible bound to the variance of Z_{DR} for such circular

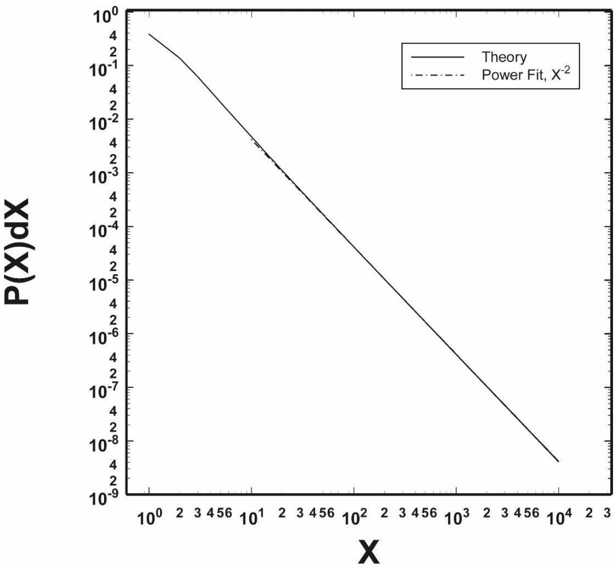

FIG. 1. The theoretical pdf of the normalized differential reflectivity X for Rayleigh statistics as discussed in the text. Note the power-law tail of the distribution which makes moments such as the mean and variance nonconvergent over [0, ∞].

Gaussian models can also be computed (Schultz and Kostinski 1997).

As useful as these results have been, however, it is becoming increasingly evident that, as a radar scans, conditions are not statistically stationary often in important meteorological settings. Rather, most precipitation is statistically heterogeneous so that radar signals do not satisfy the conditions required for Rayleigh statistics (Jameson and Kostinski 1996; Jameson 2008). That is, non-Rayleigh signal statistics are likely not as rare as previously assumed. Even in situations of statistical homogeneity, conditions may fluctuate significantly largely because of the spatial clustering of the scatterers [for a discussion see Kostinski and Jameson (1997) and Jameson and Kostinski (2000)]. When that is the case, the central limit theorem, required to produce Rayleigh statistics, is violated (Jameson and Kostinski 1996, 1999).

In light of this, it is reasonable to return to a reconsideration of the statistics of Z_{DR} but this time within the new framework of non-Rayleigh signal statistics. While the theoretical calculations performed here are for statistically homogeneous situations, we do consider some real data as well. Consequently, since real data is usually likely statistically heterogeneous, we also consider those cases when the statistical heterogeneity are consistent with the description of Jameson (2007). As argued in appendix A, even in such statistically heterogeneous cases, the statistics of the fluctuations (the standard deviation), are largely (i.e., about 80%–100% as discussed in appendix A) determined by clustering.

So how can we compare cases when clustering is absent with those when it is present? First of all, as just noted, the variance cannot be used since it does not exist for (1). Nor can we use the theoretical approach of Bringi et al. (1983) using the statistics of an estimator since the $\mathcal{R}$ and Q components are no longer Gaussian when there is clustering (shown below). This means that, unlike for Rayleigh statistics, a closed-form analytic solution corresponding to non-Rayleigh statistics is not possible. Instead, we must turn to Monte Carlo calculations as discussed in section 3 and in appendix B. Like Bringi et al. (1983) we then average over a certain number of samples of the intensities at the two polarizations to estimate Z_{DR}. We then compute the standard deviation and compare it to the Rayleigh standard deviation given by (7) in Bringi et al. (1983). As shown later, this ratio of the clustered to Rayleigh standard deviation does not depend upon the number of independent samples used to calculate an estimate of Z_{DR}. However, before doing all of this, we first briefly review clustering and present a framework used in subsequent calculations.

2. On the distribution of intensities in a clustered environment and the role of statistical heterogeneity

A detailed discussion of one derivation of the distribution of radar intensities in a clustered environment may be found in Jameson (2005). Here, the relevant points are presented with regard to this paper before proceeding with the calculations. (Some further discussions relevant to the interpretation of actual measurements in statistically heterogeneous conditions are given in appendix A with an example presented in section 4.)

The spatial variability of precipitation is widely acknowledged. How does this affect radar observations? Radars measure the reflectivity factor (Z) derived from observations of backscattered intensities (I) in sample volumes. For stationary antennas, the signal fluctuations are determined by random coherent summations of the electromagnetic waves reflected by the individual particles as described by Rayleigh statistics. The variance then equals the mean squared (i.e., the relative dispersion of the intensity, $\sigma_I/\bar{I}$, is unity for the exponential distribution of intensities arising from Rayleigh statistics). For scanning antennas, however, additional deviations occur as clustering and statistical heterogeneities extend the tail of the distributions of the relative dispersion.

That is, Rayleigh statistics are based upon the central limit theorem applied to each of the two components of the complex amplitude when conditions are "near" statistical stationarity. Jameson and Kostinski (1996) explored the meaning of near stationarity and concluded, for the simple case of drops of one size, that whenever the number of drops within the beam fluctuated from sample to sample by more than about 15% of the mean, non-Rayleigh effects could be detected. Why? Because now the measurement not only depends upon the constructive and destructive interference of the waves scattered off all the drops, it also depends upon the doubly stochastic nature of the process in which the number of scatterers themselves becomes a random variable largely because of the motion of the observation volume between successive radar samples. [A more complete discussion of the origin of non-Rayleigh signal statistics may be found in Jameson and Kostinski (1996).] What causes the number of drops to vary?

A central source of drop concentration fluctuations in a moving radar sample volume is clustering, which is the enhanced concentration (and dilution) of particles associated with increased (decreased) correlations of scatterers in neighboring volumes. That is, the number of drops in neighboring volumes is not statistically independent.

An important variable for quantitatively describing the clustering is the clustering index CI, defined by

$$\mathrm{CI} = \frac{\overline{\delta N^2}}{\overline{N}^2} - \frac{1}{\overline{N}}, \qquad (2)$$

where $\overline{N}$ and $\overline{\delta N^2}$ are the mean and variance of the number of scatterers from sample volume to sample volume members of an ensemble of such samples. Thus, when there is no clustering, $\overline{\delta N^2} = \overline{N}$ and $\mathrm{CI} \to 0$. Because the clustering intensity goes as $\overline{\delta N^2}/\overline{N}^2 - 1/\overline{N}$, the second term often quickly becomes negligible for most realistic radar sample volumes so that $\mathrm{CI} \to (\overline{\delta N^2}/\overline{N}^2)$.

As previously mentioned, another source of drop concentration variability is statistical heterogeneity, the result of changing conditions such that the statistics of the observations depend upon the location of the measurements. Consequently, radar intensity measurements exhibit increased variance (and relative dispersion) because of the bunching of drops produced by correlation (i.e., clustering) and because of systematic changes in the observed longer-term or larger-scale mean values of drop concentrations associated with statistical heterogeneity. As antennas scan during sampling, then, radars "see" the effects of all of this variability (clustering in conjunction with heterogeneity) in addition to the usual Rayleigh signal coherency fluctuations.

As the absence of any discussion in statistics textbooks indicates, the subject of statistical heterogeneity has been a difficult problem to address particularly with any generality. Recently, however, Jameson (2007) showed that statistically heterogeneous rain can apparently be decomposed into a half dozen or so statistically homogeneous components (i.e., steady drop size distributions). The statistically heterogeneous rain event, then, can be described (and, indeed, reconstituted) using linear, weighted combinations of these different statistically homogeneous components. In so far as these results are generally applicable, it then provides insight into our meteorological interpretation of non-Rayleigh signal measurements in statistically heterogeneous conditions. Using the conservative estimate discussed in appendix A and expression (A8) in Jameson (2008), it is found that

$$\frac{\sigma^2(Z)}{E^2(Z)} \cong \frac{5}{4}\sum_i \mathrm{CI}_i\left[\frac{w_i E_i(Z)}{\langle E(Z)\rangle}\right]^2 \cong \overline{\mathrm{CI}}\,\frac{5}{4}\sum_i\left[\frac{w_i E_i(Z)}{\langle E(Z)\rangle}\right]^2, \qquad (3)$$

where E denotes the expected value, w_i are the weighting functions corresponding to each of the statistically homogeneous components, $\langle\,\rangle$ denotes an ensemble average, and the "overbar" denotes a weighted average.

Thus, the squared relative dispersion of Z is determined by the summation over the clustering indices for each statistically homogeneous component of the statistically heterogeneous rain multiplied by the square of the fractional contribution each component mean $E_i(Z)$ makes to the overall average Z times a factor. Hence, one may think of the second multiplicative term as weighting factors representing the effect of statistical heterogeneities. Thus, for most radar measurements we conclude that in statistically heterogeneous rain 80%–100% of $\sigma^2(Z)/E^2(Z)$ provides a measure of a weighted average clustering index, while in statistically homogeneous rain the relative dispersion of Z provides a direct estimate of the clustering index.

It is also important to note that in (3), Z is the *intrinsic* radar reflectivity factor (without signal fluctuations) and not the measured Z. Since Rayleigh signal coherency fluctuations are present in actual observations, it is shown in Jameson (2008) that the observed and intrinsic relative dispersions are related by

$$\left[\frac{\sigma^2(Z)}{E^2(Z)}\right]_{\mathrm{Observed}} = \left[\frac{\sigma^2(Z)}{E^2(Z)}\right]_{\mathrm{Intrinsic}} + 1 = \mathrm{CI} + 1, \qquad (4)$$

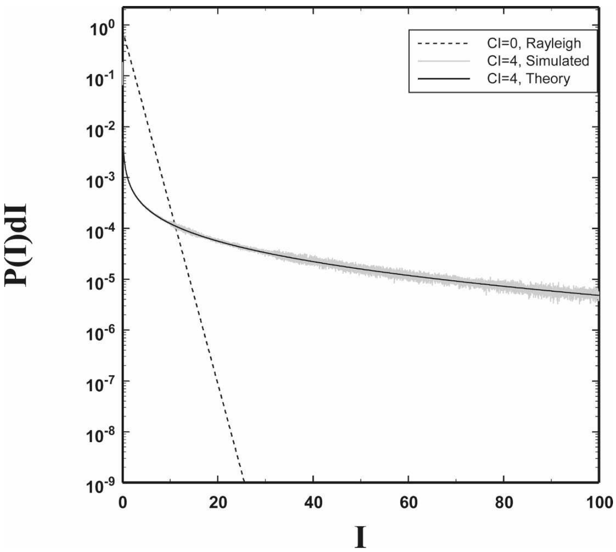

FIG. 2. Frequency distributions of the intensity normalized to the mean $\bar{I}$ for the nonclustered (Rayleigh) and clustered case as discussed in the text. Note the reliability of the Monte Carlo simulated values as indicated by the agreement between theory and simulation.

where unity appears because of the presence of Rayleigh signal fluctuations. Moreover, when there are a sufficient number of statistically independent samples, CI can then be reliably extracted from observations (Jameson 2008) by first subtracting unity from the measured $\sigma_z^2/\overline{Z}^2$ and then multiplying by a factor between 80%–100%.

3. Results from Monte Carlo simulations of signal statistics in a clustered environment

In this section we first sketch out the procedure and then discuss some results. First, a random, uncorrelated 10 million element time series is generated for the BesselK distribution (Jameson 2005) for CI = 2, 4, and 6 over the range $0 \leq I/\bar{I} \leq 200$ as discussed in appendix B. An example of the simulated frequency distribution is compared with theory in Fig. 2. Clearly, the simulations work quite well. The difference between the Rayleigh statistics (CI = 0) exponential distribution and the clustered non-Rayleigh statistics is strikingly apparent. The long tails of the clustered distributions indicate that the standard deviations of the intensities should be significantly enhanced beyond that for the Rayleigh exponential distribution of intensities. One initially suspects, then, that this could potentially have an effect on the variance of the differential reflectivity.

To explore this, we next generate a pair of correlated time series of BesselK distributed intensities, one for horizontal and one for vertical polarizations having a ρ_{HV} ranging from approximately 0.2 to 0.995. These correlated intensity time series can be used in two ways. First, Fig. 2 suggests that the distributions of complex amplitude component, $\mathcal{R}$ and Q, may no longer be Gaussian. Indeed, this is the case as Fig. 3a confirms. The distribution is definitely non-Gaussian with tails that far exceed those for the normal distribution for the same mean intensity. While somewhat like a Lorentz distribution in appearance, a closer inspection (not shown) reveals that the distribution actually has longer

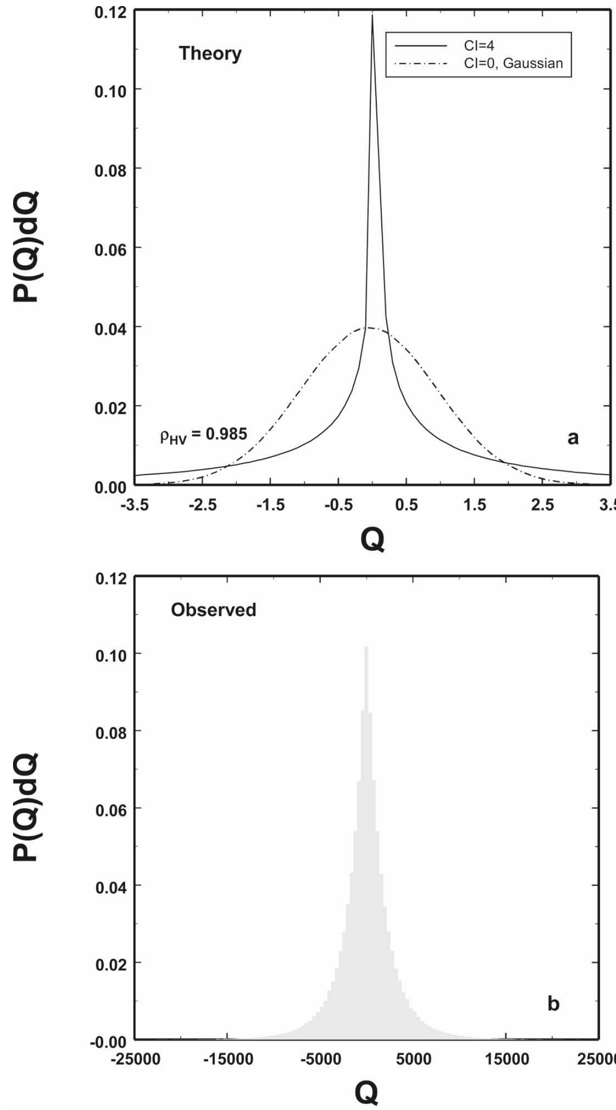

FIG. 3. (a) Frequency distributions for the Q component of the complex amplitude for Rayleigh signal statistics (the Gaussian distribution) and simulated clustered case for the same mean I as discussed in the text. (b) An example of the distribution of Q measured by a scanning radar.

and more significant tails than does a Lorentz pdf. In fact, it can be shown that, for the BesselK distribution of intensities, the distribution for the Q component is given by

$$P(Q) = \sqrt{\frac{2}{\pi}\frac{1}{\Gamma(1/\mathrm{CI})}}\left(\frac{1}{B}\right)^{(1/4)(1+2/\mathrm{CI})}\left(\frac{1}{Q^2}\right)^{(1/4)(1-2/\mathrm{CI})}\mathrm{BesselK}\left[\frac{1}{\mathrm{CI}}-\frac{1}{2},\frac{2\sqrt{Q^2}}{\sqrt{B}}\right], \tag{5}$$

where CI is the clustering index, BesselK is the modified Bessel function of the first kind,

$$B = \frac{\mathrm{rms}(Q)}{\sqrt{(\mathrm{CI})^2 + \mathrm{CI}}},$$

and rms denotes the root-mean-square. The same distribution also holds for $\mathcal{R}$. Such distributions (Fig. 3a) appear in observations as well (Fig. 3b), giving us some confidence in this approach.

Second, the statistics of Z_{DR} can then be calculated. Figure 4 shows an observed distribution of Z_{DR} compared to the appropriate distributions assuming Rayleigh and clustering statistics. These data are from dual-polarization time series measurements of the $\mathcal{R}$ and Q components at alternating polarizations from a stationary antenna over several seconds yielding around 80 000 independent samples over the same locations. Non-Rayleigh effects are then observed by combining observations along the beam. This is analogous to looking at data from a single range bin as it scans rapidly multiple times through the same set of heterogeneous conditions. Here the parameter X is the ensemble observed value normalized by the expected Z_{DR}, as defined previously. The difference between these two frequency distributions of X for the measured $\rho_{\mathrm{HV}} = 0.970$ are fairly typical. We note the broadening of the distribution in the clustered environment as compared to that corresponding to Rayleigh statistics (CI = 0). The likely source of this enhanced broadening is the extended tails of $\mathcal{R}$ and Q (Fig. 3) as exhibited by the extended tails of the intensity distribution (Fig. 2). As with the Rayleigh statistics, ρ_{HV} acts to significantly reduce the spread in the distribution of Z_{DR}, but in the clustered environment, that is still insufficient to return the variance of Z_{DR} to Rayleigh values.

Now in Fig. 4, there is no averaging; that is, the curves are the frequency distributions of statistically independent pulse to pulse values of X. In reality something like 5–10 independent samples will provide an ensemble average so that both distributions in Fig. 4 would be narrower. Regardless, however, what is important is that the standard deviation in the clustered environment will always be greater than the value computed assuming Rayleigh statistics. This is illustrated in greater detail in Fig. 5.

Clearly, in all situations, clustering enhances the standard deviation Z_{DR} beyond what would be calculated assuming Rayleigh statistics, and the greater the clustering index, the greater this effect even when $\rho_{\mathrm{HV}} \geq 0.99$. The symbol in the inset frame in Fig. 5 represents the observed data in Fig. 4 for the CI = $\sigma_z^2/\overline{Z}^2$; that is, the CI is 100% of the relative dispersion of Z (see appendix A). The agreement is remarkable despite the observation that the distributions of Z_{DR} (and the distributions of the intensities) differ somewhat.

As Fig. 5 illustrates, ρ_{HV} plays a crucial role in the accuracy of the Z_{DR} measurements. It is important, then, to get ρ_{HV} as close as possible to unity. It has been suggested (Chandrasekar et al. 1994) that this might be done using simultaneous horizontally and vertically polarized transmissions (the so-called 45° transmissions) in order to minimize decorrelation caused by particle reshuffling during one interpulse period (a problem in the cases of transmissions of alternating polarization from pulse to pulse). Despite this approach, however, there will always be decorrelation for several reasons including the meteorologically intrinsic differences of the scatterers viewed at the two different polarizations (Jameson 1989) and the differences in the two radar beam patterns at each polarization.

Of course, it must be remembered that clustering will have no effect on the measurements if the antenna remains stationary unless data in different data bins are combined (as in Fig. 4). Thus, one way around the enhanced uncertainty induced by clustering is to take observations in azimuthal incremental steps keeping the antenna stationary at each step during sampling or simply to rotate the antenna very slowly. These alternatives, however, are not consistent with the rapid scan often desired for greater temporal resolution with the potential exception of electronically steered, dual-polarized phased array meteorological radars if that should ever come to pass. On the other hand, for some meteorological purposes, non-Rayleigh signals may actually be useful (Jameson 2008). Some investigators may, in fact, wish to enhance rather than to suppress non-Rayleigh signals as a way to explore subbeam-scale variability.

Another important point to remember is that the ratios in Fig. 5 will largely be independent of the number

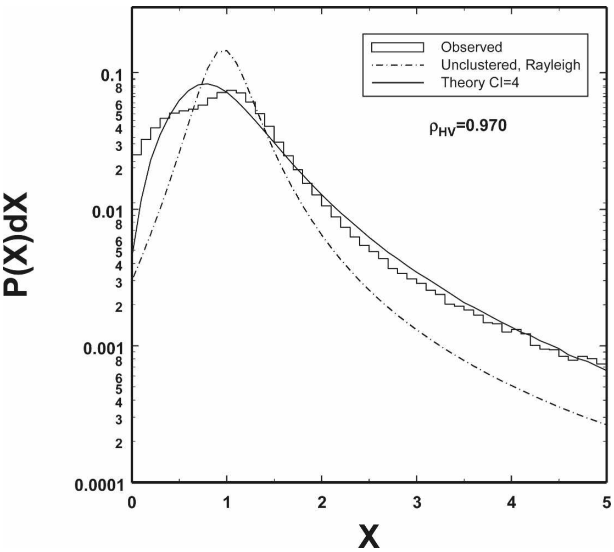

FIG. 4. Frequency distributions of the normalized differential reflectivity for ρ_{HV} = 0.970. Note the significant broadening because of clustering. The theoretical curve for CI = 4 shows the same general structure as the radar observations.

of independent samples m. This occurs because both $\sigma_{\mathrm{Clustered}}$ and $\sigma_{\mathrm{Rayleigh}}$ essentially depend on the number of independent samples in the estimate as $1/\sqrt{m}$ (Fig. 6a) so that the ratio of the clustered to Rayleigh standard deviations are independent of m (Fig. 6b). Thus, for example, if one calculates that $\sigma_{Z_{\mathrm{DR}}} \approx 0.2$ dB, (or 1.047 on a linear scale) as Gorgucci et al. (2006) suggest, the standard deviation may actually be, say, 0.4 dB (1.097) to 1.2 dB (1.230) or larger depending upon CI and ρ_{HV} at the location of the observation.

The objective of this work, then, is simply to suggest using reasonable caution by pointing out the potential effect of clustering on measurements of Z_{DR} and to

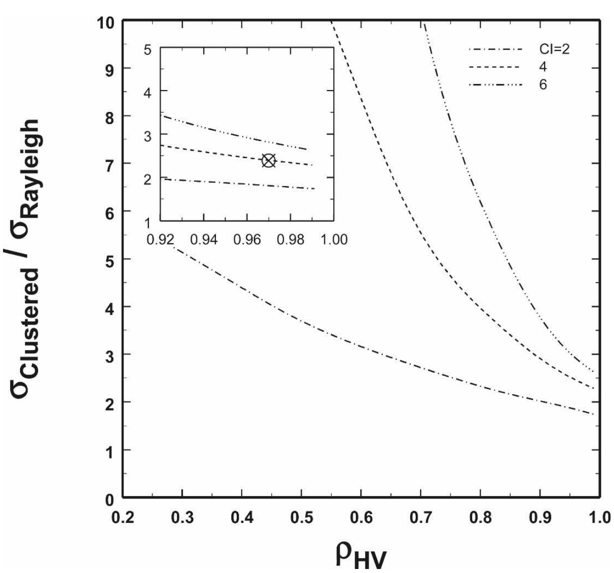

FIG. 5. The multiplicative factor of clustering on the standard deviation (uncertainty) of the differential reflectivity calculated under the usual assumption of Rayleigh statistics. The effect occurs for all at high correlations ρ_{HV} even those approaching unity. The impact of clustering also increases with increasing CI. The symbol in the inset corresponds to the data in the previous figure as discussed in the text.

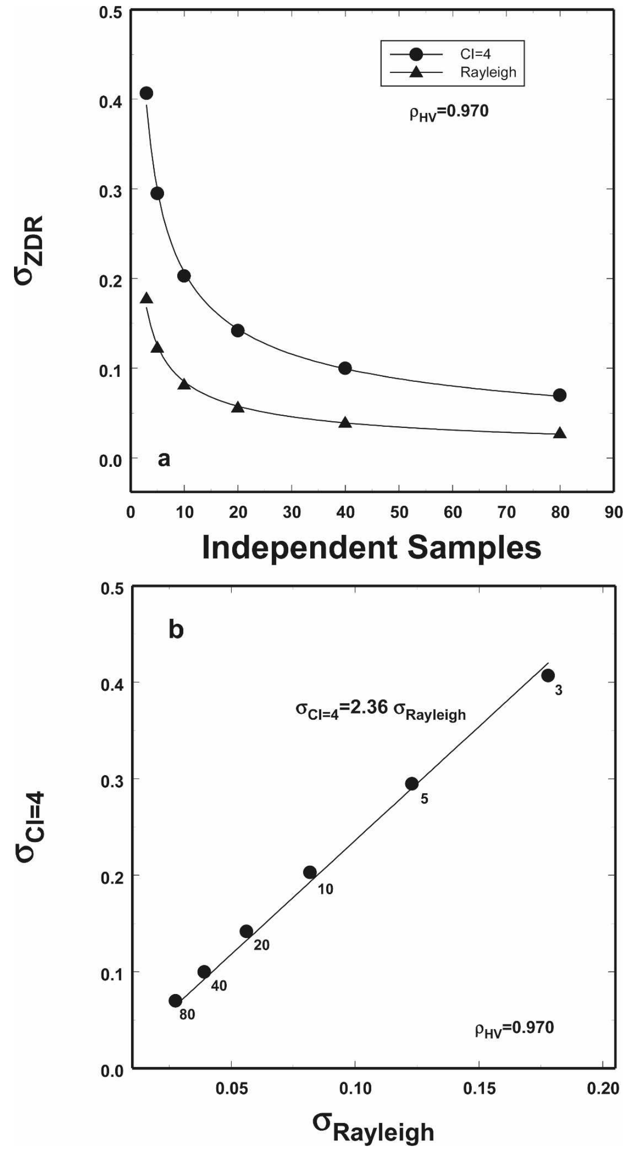

FIG. 6. (a) Standard deviations for clustered and Rayleigh cases for average powers ratio estimator of normalized Z_{DR} on a linear scale. Both exhibit nearly inverse square root relationships to the number of independent samples. (b) An illustration showing that the ratio of the clustered to Rayleigh standard deviations does not depend upon the number of independent samples (numbers in the figure) used to estimate Z_{DR}.

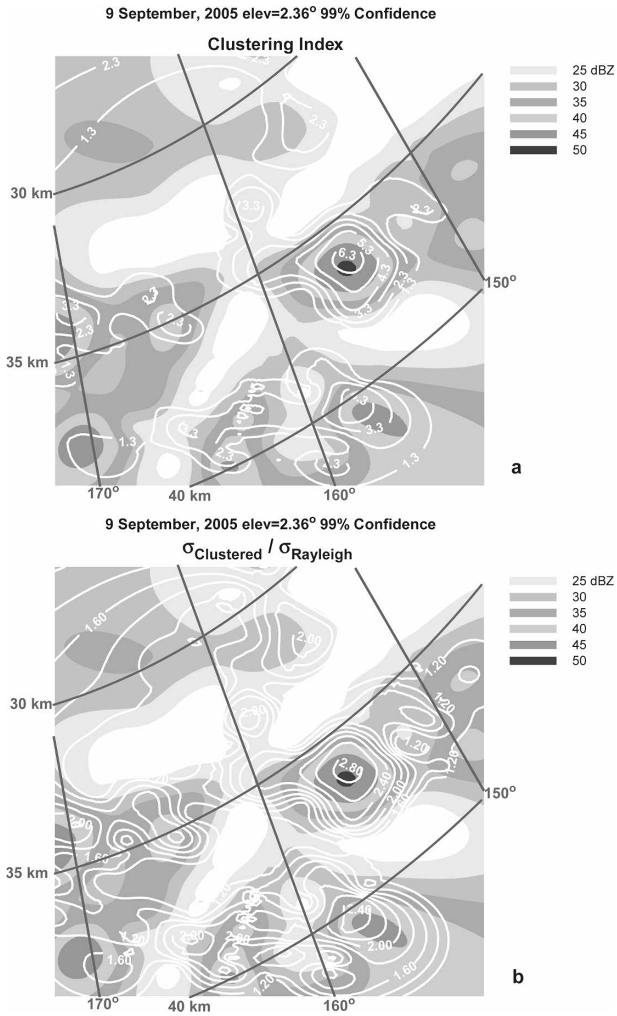

FIG. 7. (a) Shaded contour plots of the radar reflectivity factor, Z, in units of dBZ with overlays of contours of the CI at the 99% confidence level as functions of the radar range and azimuth (after Jameson 2008, his Fig. 4). CI is estimated assuming it is 80% of $\sigma_Z^2/\overline{Z}^2$ as discussed in appendix A and the text. Note that the radials and arcs are not orthogonal since the scaling in the vertical and horizontal directions differ in this plot. (b) The multiplicative factor of the clustered to the traditional, nonclustered (Rayleigh) standard deviation contoured beginning at 1.2 in steps of 0.2 and overlaid on the shaded contour plot of Z using the clustered indices in (a) as discussed in the text.

leave any rational scientist at least thinking about just how much such clustering may affect the future operational accuracy of some polarization rainfall algorithms. At the very least, values as low as Gorgucci et al. (2006) suggest seem highly improbable much of the time as illustrated later in Fig. 7 except, perhaps, in less interesting meteorological conditions.

However, with our current nearly nonexistent measurements of the clustering index it is impossible to generalize the extent of the effect of clustering on radar measurements except to note that, regardless of radars, clustering appears to be ubiquitous throughout most precipitation. When and where that will be significant remains to be determined through extensive observations. However, in the next section, one example from a rare set of observations of CI is presented. While its relevance should not be viewed with any sweeping generality, of course, it does raise valid concerns and suggests the importance of gathering further observations of the clustering index in a number of different meteorological settings.

4. A sample observation

The azimuthal-range scan considered in this section was measured on 9 September 2005. Details may be found in Jameson (2008), but some of the relevant information is repeated here as well. These data were kindly provided by Dave Brunkow of the University of Chicago–Illinois State Water Survey (CHILL) National Radar Facility at Colorado State University in Fort Collins, Colorado. In both cases the observations were made in rain by the CHILL radar that has a nominal 1° beam and wavelength of 10 cm.

In each scan, over a set of 150-m range bins and over several degrees of azimuth, the $\mathcal{R}$ and Q data were given for each pulse. This permitted a detailed inspection of data quality (e.g., the $\mathcal{R}$ and Q were found to be completely balanced). It was also then possible to compute the intensity ($\mathcal{R}^2 + Q^2$) for each pulse and range bin. Data over a number of pulses and range bins were then used to estimate $\overline{I^2}$ and $\overline{I}^2$ and, therefore, the variance $\sigma^2(I)$. The number of data bins combined was selected to preserve as high a spatial resolution as possible while still yielding a reasonable number of statistically independent samples. The number of statistically independent samples was estimated assuming that there were 5 independent samples per 1° of azimuth and that the range bins along each azimuthal radial separated by 1° were also statistically independent. This means that the estimate of the number of independent samples for each estimate of the variance and mean of I were, on average, on the order of 13 in this example. As explained in Jameson (2008), the corresponding 99% confidence threshold that the observed relative dispersions (squared) of Z_H were not Rayleigh in origin was found to be 2.24 (i.e., a CI of 1.24). That is, any CI greater than 1.24 were 99% likely *not* due to Rayleigh fluctuations, or to put it more positively, they were 99% *likely* due to clustering (weighted by statistical heterogeneity, of course).

To see where clustering is occurring, the shaded contours of radar reflectivity Z_H are shown in Fig. 7a adapted from Jameson (2008) along with the contours of the clustering index in unit steps beginning with the modest CI = 1.3. In this example, the most significant values of the clustering index are often found in or near the highest radar reflectivity factors such as in the core of a small, convective shower at about 156° azimuth and 36-km range (156°, 36 km). This need not always be the case, however (Fig. 5 of Jameson 2008).

To get an idea of what this means with respect to the level of uncertainty in Z_{DR} we assume that $\rho_{HV} = 0.98$. From the numbers used to generate Fig. 5, it is found that the ratio of $\sigma_{Clustered}/\sigma_{Rayleigh}$ goes as $(1.00 + 1.083CI)^{1/2}$ at this ρ_{HV}. Using this expression, the clustering indices are converted into the ratio of the standard deviations. These are plotted in Fig. 7b in steps of 0.2 beginning at 1.2. Clearly, there are areas over which the uncertainty is significantly greater than Rayleigh statistics would imply. So, for example, near the core of the shower in Fig. 7, one would have to multiply $\sigma_{Rayleigh}$ by a factor of more than 2.8, yielding an uncertainty of almost 1 dB if the estimated uncertainty based on a Rayleigh assumption had been 0.4 dB. Thus, in this case the uncertainty is greatest just where accurate estimates of, say, the rainfall rate are most desirable, namely near the core of the most intense shower.

5. Summary and conclusions

Polarization measurements are poised to play an increasingly important role in future operational radar systems. While polarization observations can contribute in many ways, one of the most quantitative applications will be to radar rainfall measurement. In particular, many studies have suggested that polarization rainfall estimates, for example, can significantly improve such estimates relative to those derived using the standard radar reflectivity–rainfall rate (*Z*–*R*) relations in current use.

A key parameter in all of these polarization–rainfall algorithms is the differential reflectivity (Z_{DR}), the ratio of the radar reflectivity factor measured using horizontally polarized waves to that observed using vertical polarization (Z_H/Z_V). As with any scientific variable, it is important to understand the statistics associated with Z_{DR}. This was first addressed extensively by Bringi et al. (1983) and in subsequent studies under the assumption that Rayleigh signal statistics, as expounded for radar by Marshall and Hitschfeld (1953), apply. This has been the solitary approach to radar signal statistics since the beginning days of radar meteorology.

Rayleigh statistics, however, are based upon two essential assumptions, namely that conditions are statistically stationary during sampling and that each particle scatters independently. It is becoming increasingly evident, however, that as a radar scans and in important meteorological settings, conditions are not statistically stationary in time (i.e., they are statistically heterogeneous across the sampling space) so that often, radar signals do not satisfy the conditions required for Rayleigh statistics (Jameson 2008). Even in those rare situations of statistical homogeneity, conditions may fluctuate significantly largely because of clustering or bunching of the scatterers (Kostinski and Jameson 1997; Jameson and Kostinski 2000) so that the central limit theorem, required to produce Rayleigh statistics, is violated (Jameson and Kostinski 1996, 1999). Since it appears meteorologically that the clustering of particles is nearly ubiquitous, non-Rayleigh signal statistics can no longer be attributed to unusual meteorological conditions.

Therefore, in this study we return to a reconsideration of the statistics of Z_{DR} but this time within the new framework of non-Rayleigh signal statistics. Here, as discussed in the text [and in so far as the results in Jameson (2007) apply], even in conditions of statistical heterogeneity, clustering is largely responsible for the statistics of the fluctuations (see appendix A and Figs. 4 and 5), although other terms due to statistical heterogeneity may also contribute at times (appendix A). Consequently, as has been done in all previous studies, in the theoretical analysis we consider only statistically homogeneous but clustered conditions remembering that our conclusions are likely applicable in a substantial degree to those statistically heterogeneous cases consistent with the results in Jameson (2007) as argued in appendix A.

For reasons given in Jameson (2005), the BesselK function is used as the probability density function (pdf) of radar intensity measurements in a clustered environment. This has the advantage that the clustering index is explicit in the intensity distribution. However, because of the correlation, ρ_{HV}, between Z_H and Z_V, a direct analytic solution for the statistics of Z_{DR} is not possible so that Monte Carlo calculations (appendix B) are used to study the effect of clustering on the standard deviation of Z_{DR} using the estimator of averaged powers [Bringi et al. 1983, their Eq. (4)].

It is found that $\sigma_{Z_{DR}}$ is significantly enhanced by clustering to a degree that depends upon the clustering index and ρ_{HV}. It is noteworthy, however, that, regardless of how small the Rayleigh standard deviation is reduced by increasing the number of independent samples *m*, clustering still magnifies that Rayleigh standard deviation by an amount that is independent of *m*. This effect is present even for ρ_{HV} approaching unity and over the range of CI values likely to be found in most real observations.

As pointed out in Jameson (2008), most current radar observations of reflectivity factors alone are unlikely to be too affected by non-Rayleigh signal statistics because of the small number of independent samples used to form estimates. However, that is not the case for more sophisticated observations such as the differential reflectivity involving ratios of such parameters as Fig. 6 demonstrates. All current observations of Z_{DR} are likely affected to varying degrees. Using the methodology described here and in Jameson (2008), an example in convective showers suggests that this amplification of $\sigma_{Z_{\mathrm{DR}}}$ by non-Rayleigh signals should be taken into account in future implementations of polarization radar rainfall estimation algorithms. At the very least, it is time to begin to document the prevalence and magnitude of clustering and its subsequent generation of non-Rayleigh signal statistics in a wide variety of meteorological settings if we are going to embrace with confidence more quantitative applications of radar observations.

Acknowledgments. This work was supported by the National Science Foundation (NSF) under Grants ATM05–31996 (AJ) and ATM05–54670 (AK).

APPENDIX A

On the Contribution of Clustering to the Relative Dispersion of the Radar Reflectivity Factor in Statistically Heterogeneous Environments

In previous work (Jameson 2008) the relationship between the relative dispersion of the radar reflectivity factor and the clustering index in a statistically heterogeneous environment was developed under the assumption that the weighting functions of the statistically homogeneous components of the heterogeneous data were represented by a Dirac distribution. Although nothing is yet known about those distributions, it seems likely that they will usually have some breadth expressed as a variance. It is shown below that when that is the case the variance (and, hence, the relative dispersion) of the radar reflectivity factor is enhanced by factors other than just the clustering, although that remains the predominant influence.

Specifically, we imagine an entire set of data having associated sets of weighting functions $\{w_i\}$ and realizations of M components of the radar reflectivity factors $\{Z_i\}$. For one realization from this ensemble of observations $Z' = \sum_{i=1}^{M} w_i' Z_i'$ it then follows that

$$\overline{Z} = \langle Z' \rangle = \sum_{i=1}^{M} \langle w_i' Z_i' \rangle = \sum_{i=1}^{M} \langle w_i' \rangle \langle Z_i' \rangle = \sum_{i=1}^{M} w_i Z_i, \text{ while} \tag{A1}$$

$$\begin{aligned}\overline{Z^2} = \langle Z'^2 \rangle &= \left\langle \left(\sum_{i=1}^{M} w_i' Z_i' \right)^2 \right\rangle = \left\langle \sum_{i=1}^{M} w_i'^2 Z_i'^2 \right\rangle + \left\langle \sum_{i \neq}^{M} \sum_{j}^{M} w_i' w_j' Z_i' Z_j' \right\rangle \\ &= \sum_{i=1}^{M} \langle w_i'^2 \rangle \langle Z_i'^2 \rangle + \sum_{i \neq}^{M} \sum_{j}^{M} \langle w_i' w_j' \rangle \langle Z_i' Z_j' \rangle, \quad \text{so that}\end{aligned} \tag{A2}$$

$$\sigma_Z^2 = \overline{Z^2} - \overline{Z}^2 = \sum_{i=1}^{M} \sigma_{w_i}^2 \sigma_{Z_i}^2 + \sum_{i=1}^{M} \sigma_{w_i}^2 Z_i^2 + \sum_{i=1}^{M} w_i^2 \sigma_{Z_i}^2 + \sum_{i \neq}^{M} \sum_{j}^{M} Z_i Z_j \operatorname{cov}(w_i w_j), \tag{A3}$$

where it is assumed that the Z_i are statistically independent. In contrast to the derivation in Jameson (2008) additional terms are now included to reflect the complete contribution from the statistical heterogeneity. This means that the observed relative dispersion of Z will be enhanced beyond that due solely to clustering. Let us estimate by how much.

Give our current state of knowledge, it is now assumed that the ws are statistically independent so that (A3) becomes

$$\sigma_Z^2 = \overline{Z^2} - \overline{Z}^2 = \sum_{i=1}^{M} \sigma_{w_i}^2 \sigma_{Z_i}^2 + \sum_{i=1}^{M} \sigma_{w_i}^2 Z_i^2 + \sum_{i=1}^{M} w_i^2 \sigma_{Z_i}^2. \tag{A4}$$

Unfortunately, at present nothing is known about the pdf of w. Consistent with this ignorance, the assumption is also made that the distributions are uniform over [0, 1] for all w_i. In that case $\sigma_{w_i}^2 = 1/3w_i^2$ so that (A4) reduces to

$$\sigma_Z^2 = \overline{Z^2} - \overline{Z}^2 = \frac{1}{3} \sum_{i=1}^{M} w_i^2 Z_i^2 + \frac{4}{3} \sum_{i=1}^{M} w_i^2 \sigma_{Z_i}^2. \tag{A5}$$

Now, because of different meteorological conditions even within one set of observations, it is unlikely that the Z_i are uniformly distributed. Since Z_i^2 is of the order $\sigma_{Z_i}^2$, to a reasonable approximation we can then write

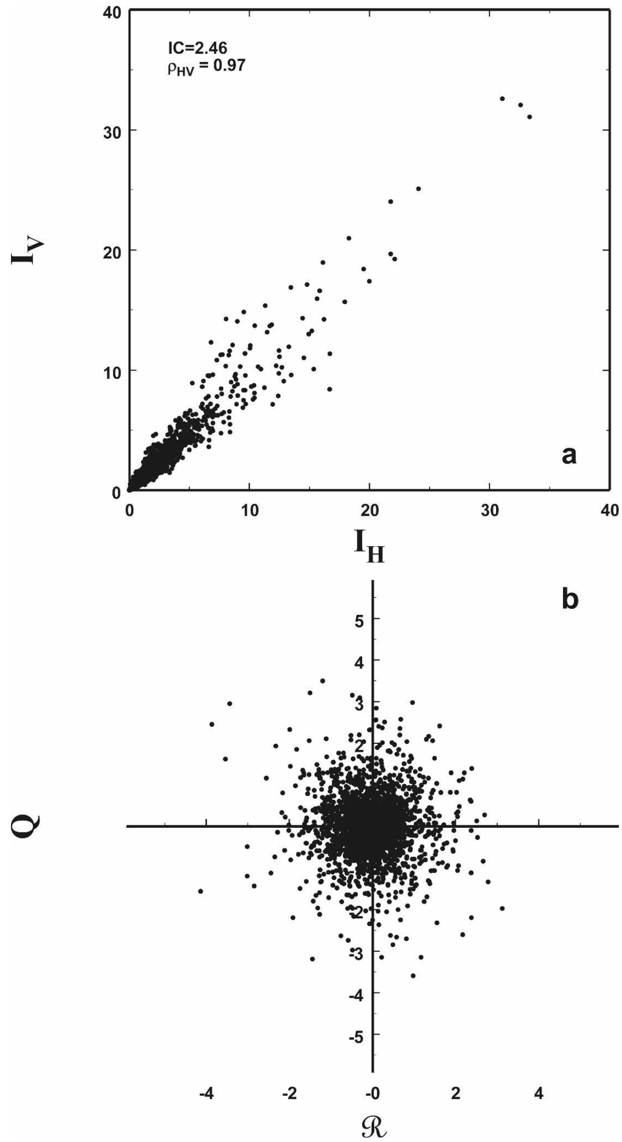

FIG. B1. (a) An example from the Monte Carlo procedure described in appendix B illustrating that the correlation between the intensities normalized to their means at vertical (I_V) and horizontal (I_H) polarizations is preserved during simulation. (b) Scatterplot of the $\mathcal{R}$ and Q components of the complex amplitude corresponding to one case from the Monte Carlo simulation procedure described in appendix B. The two components are statistically independent as they should be.

$$\sigma_Z^2 = \overline{Z^2} - \overline{Z}^2 \approx \frac{5}{3} \sum_{i=1}^{M} w_i^2 \sigma_{Z_i}^2. \tag{A6}$$

Consequently, it is still expected that the clustering index will be a minimum of at least 60% of the intrinsic relative dispersion, and it seems far more likely (using more reasonable pdfs for w_i) that the minimum is more likely closer to 80%–100%. Furthermore, as the pdf of w_i narrows, this percentage will increase until it is 100% for the Dirac distribution of w_i as for the data in Figs. 4 and 5. However, to remain conservative, in this work it is assumed that the CI is represented by 80% of the intrinsic relative dispersion, realizing that in Fig. 7 that assumption may well be an underestimate.

APPENDIX B

The Monte Carlo Generation of the $\mathcal{R}$, Q Components and Correlated Intensities in a Clustered Environment

A random sequence of draws from the BesselK distribution of intensities corresponding to a clustering indices CI = 2, 4, and 6 are generated using the copula transformation method (Genest and Mackay 1986; Nelsen 1999; Frees and Valdez 1998) using correlated Gaussian distributions and a uniform random number generator (Fox et al. 1988). First, corresponding to a particular ρ_{HV}, two correlated time series of zero mean, Gaussian distributions with unit variance are generated. Using the inverse accumulated density function (CDF) for the Gaussian, these time series are then converted into correlated time series of uniformly distributed variables. Next, using the CDF of BesselK pdf, found through fitting procedures, one can generate an inverse function to transform the uniform random variables into BesselK random variables. The correlated intensities at both horizontal and vertical polarizations as well as Z_{DR} can then be computed as illustrated in Fig. B1a.

Second, the $\mathcal{R}$, Q at both polarizations can also be computed by taking the square root of the just correlated intensities (I_H and I_V) and then picking uniformly uncorrelated random phases (ϕ) so that $\mathcal{R}_{H,V} = (A_H, A_V)\cos(\phi_i)$ and $Q = (A_H, A_V)\sin(\phi_i)$. Consequently, the random phases between $\mathcal{R}$ and Q are produced (Fig. B1b), while the intensities at the two polarizations remain correlated.

REFERENCES

Bringi, V. N., T. A. Seliga, and S. M. Cherry, 1983: Statistical properties of the dual-polarization differential reflectivity (ZDR) radar signal. *IEEE Trans. Geosci. Remote Sens.,* **21,** 215–220.

Chandrasekar, V., J. Hubbert, and V. N. Bringi, 1994: Analysis and interpretation of dual-polarized radar measurements at +45° and −45° linear polarization states. *J. Atmos. Oceanic Technol.,* **11,** 323–336.

Fox, R. F., I. R. Gatland, R. Roy, and G. Vermuri, 1988: Fast, accurate algorithm for numerical simulation of exponentially correlated colored noise. *Phys. Rev.,* **38,** 5938–5940.

Frees, E., and E. Valdez, 1998: Understanding relationships using copulas. *North Amer. Actuarial J.,* **1,** 1–25.

Genest, C., and J. Mackay, 1986: The joy of copulas: Bivariate distributions with uniform marginals. *Amer. Stat.,* **40,** 280–283.

Gorgucci, E., L. Baldini, and V. Chandrasekar, 2006: What is the shape of a raindrop? An answer from radar measurements. *J. Atmos. Sci.,* **63,** 3033–3044.

Jameson, A. R., 1989: The interpretation and meteorological application of radar backscatter amplitude ratios at linear polarizations. *J. Atmos. Oceanic Technol.,* **6,** 908–919.

——, 2005: Measuring clustering in clouds using non-Rayleigh signal statistics. *J. Atmos. Oceanic Technol.,* **22,** 490–495.

——, 2007: A new characterization of rain and clouds: Results from a statistical inversion of count data. *J. Atmos. Sci.,* **64,** 2012–2028.

——, 2008: Radar observations of rainfall variability using non-Rayleigh signal fluctuations. *J. Appl. Meteor. Climatol.,* **47,** 607–619.

——, and A. B. Kostinski, 1996: Non-Rayleigh signal statistics caused by relative motion during measurements. *J. Appl. Meteor.,* **35,** 1846–1859.

——, and ——, 1999: Non-Rayleigh signal statistics in clustered statistically homogeneous rain. *J. Atmos. Oceanic Technol.,* **16,** 575–583.

——, and ——, 2000: Fluctuation properties of precipitation. Part VI: Observations of hyperfine clustering and drop size distribution structures in three-dimensional rain. *J. Atmos. Sci.,* **57,** 373–388.

Kostinski, A. B., and A. R. Jameson, 1997: Fluctuation properties of precipitation. Part I: On deviations of single-size counts from the Poisson distribution. *J. Atmos. Sci.,* **54,** 2174–2186.

Krishnaiah, P. R., P. Hagis Jr., and L. Steinberg, 1963: A note on the bivariate Chi distribution. *SIAM Rev.,* **5,** 140–144.

Marshall, J. S., and W. Hitschfeld, 1953: Interpretation of the fluctuating echo from randomly distributed scatterers. Part I. *Can. J. Phys.,* **31,** 962–994.

Nelsen, R. B., 1999: *An Introduction to Copulas.* Springer, 216 pp.

Rayleigh, J. W. S., 1877: *The Theory of Sound.* Vol. I. Reprint, Dover, 1945, 984 pp.

Schaffner, M. R., R. Rinehart, and R. Leonardi, 1980: Detection of nonuniformities within the radar measurement cell. Preprints, *19th Conf. on Radar Meteorology,* Miami, FL, Amer. Meteor. Soc., 256–263.

Schultz, T. J., and A. B. Kostinski, 1997: Variance bounds on the estimation of reflectivity and polarization parameters in radar meteorology. *IEEE Trans. Geosci. Remote Sens.,* **35,** 248–255.

Srivastava, R. C., and R. E. Carbone, 1971: The effect of signal fluctuations on the performance of a dual wavelength radar hail detector. University of Chicago, Department of Geophysical Sciences, Laboratory for Atmospheric Probing Rep. 20, 23 pp.